

情報시스템의 評價와 分析 (2)

金 英 男
 <KORSTIC 企劃室>

이번 號에서는 前篇에서 紹介한 몇가지 基本 概念을 中心으로 情報시스템의 費用-效果-利潤 關係를 좀더 상세히 다루고자 한다.

3. 費用-効率(性能)-利潤

情報시스템의 効率이나 性能은 여러가지 方法을 통해 多樣하게 測定해 낼 수 있다. 이들 測定 方法은 각기 獨特한 測定基準를 바탕으로 形成되어 있다.

예를 들어 遡及調査시스템의 性能測定은 다음과 같은 몇가지 測定基準를 中心으로 施行된다.

- 1) 檢索範圍
- 2) 檢索能力
- 3) 檢索의 正確性
- 4) 檢索所要時間

이들 性能測定基準는 遡及調査뿐 아니라 SDI와 같은 他 檢索시스템에도 適用될 수 있다.

情報시스템에 있어 利潤을 正確히 測定해 내기란 매우 어려운 일이다. 利潤概念 自体가 擴大된 他意味까지를 포함하고 있기 때문이다. 따라서 利潤을 現實적으로 測定해 내기 위해서는 몇가지 測定基準를 設定해야 할 것이다.

- 1) 他시스템과 비교할 때 認知할 수 있는 費用節減
- 2) 意思決定能力의 增進
- 3) 不必要한 研究의 重複과 努力消耗의 防止
- 4) 研究意慾의 鼓吹

위에 列擧한 事項들은 시스템의 利潤測定基準으로서 特別한 意味를 지니고 있다.

情報시스템에 있어 利潤은 多樣한 型態로 나타난다. 예를 들어 協會나 組合 등에서는 情報시스템의 利潤을 出版物이나 情報서비스의 販賣를 통한 收益의 概念으로 파악하고 있다. 반면에 政府機關에서는 시스템의 利潤을 測定不可能한 부분까지 擴大하여 파악하고 있다.

費用-效果-利潤은 相互 密接히 聯關되어 있기 때문에 分離하여 생각할 수는 없다. 費用效果分析은 支出된 費用에 대한 價値의 增殖, 즉 經濟性 向上을 그 目的으로 하고 있다. 시스템의 經濟性 向上은 2가지 方向으로 推進해 나갈 수 있을 것이다.

- 1) 現狀維持된 시스템効率下에서 시스템 運用 費用을 節減
- 2) 시스템 運用費用을 現狀維持하고 시스템効率을 向上

위와 같은 方向으로 經濟性 向上을 추구해 나가기 위해서는 다음과 같은 몇가지 基本段階를 거쳐 費用效果分析을 實施해 나가야 할 것이다.

- 1) 達成할 目標의 定義
- 2) 目標達成을 위한 方案確認
- 3) 각 方案에 所要되는 費用算定
- 4) 각 方案에 대한 費用面에서의 모델設定
- 5) 最適方案을 選定키 위한 基準設定

費用效果分析을 위한 基本段階를 바탕으로 情報시스템의 性能과 관계있는 몇가지 重要要素-蒐集蓄積機能, 確認處理機能, 提供機能-를 생각해 보자.

蒐集蓄積費用은 必要項目數, 豫想情報需要, 項目當 平均購入費用, 文獻의 크기 등에 의해 좌우

된다. 文獻의 크기는 蓄積費用과 複寫費用에도 영향을 미친다. 確認處理費用은 文獻의 크기, 難易度, 索引語의 크기와 形態, 索引語의 網羅度, 索引者의 專門性과 給與水準, 索引者의 個人能力, 平均調査所要時間 등에 의해 좌우된다. 確認調査機能의 性能은 再現率, 正確率, 反應時間, 利用者 投入勞力 등으로 나타낼 수 있다.

提供費用은 提供파일의 크기에 의해 決定된다. 情報파일의 크기는 平均檢索項目數, 正確度, 印刷方法, 情報提供形態 등에 차례로 영향을 미친다.

各 機能의 効率은 他機能의 効率에도 영향을 미친다. 正確率은 確認處理機能의 効率測定媒体임과 동시에 提供費用에 영향을 미치는 要素이기도 하다.

情報시스템에서 調査効率에 관계되는 要素는 무수하다.

그림 5는 情報利用者가 情報를 要求할 때부터 必要情報가 그에게 提供될 때 까지의 遡及調査 시스템 基本段階를 圖示하고 있다.

시스템은 關聯要素의 相互作用에 의해 그 性能이 떨어질 수도 있다. 우선 情報要求가 不明確하면 必要情報중 20% 정도가 檢索對象에서 제외되고, 檢索用語까지 제한시킨다면 나머지중 10%도 檢索에서 누락될 것이다. 그외에도 不適合한 調査方法에 의해 15%, 索引의 不合理性에 의해 5%, 調査後 整理作業에 의해 5% 정도의 項

目이 檢索作業에서 제외될 것이다.

結果적으로 情報要求者가 利用할 수 있는 項目은 전체의 45%에 지나지 않는다. 물론 위의 數置는 假定値에 불과하나 이러한 減殺效果에 의해 필요한 情報를 100% 檢索해 낼 수 없는 것만은 사실이다. 情報利用者의 要求를 情報相談者가 어떻게 解析하느냐에 따라 性能減殺效果가 크게 달라진다. 왜냐하면 情報相談者의 解析 여부에 따라 調査方針이 決定되기 때문이다.

그림 5는 情報檢索의 각 단계와 이 단계의 成敗를 좌우하는 각 要素를 표시하고 있다. 이제 부터 각 단계에 費用效果分析을 실제 適用시켜 보고자 한다.

4. 資料의 選定

情報시스템에서는 情報蓄積을 위한 資料選定에 現在의 可用資源을 어떻게 最適配分할 것인가에, 費用效果面에서 크게 관심을 두고 있다. 資料選定基準은 사실 情報시스템의 設計와 評價에서는 輕視되어 왔지만 費用效果面에서는 큰 비중을 차지하고 있다.

潛在 情報利用人口까지를 감안하여 모든 資料를 蒐集索引한다는 것은 不可能한 일이다. 따라서 資料는 情報서비스 利用者에게 가장 適合한 것만이 選擇되어야 할 것이다. 모든 資料를 利用度에 따라 序列化할 수 있다면 그것은 資料蒐集의 가장 理想的인 方法이 될 것이다. 이 方法

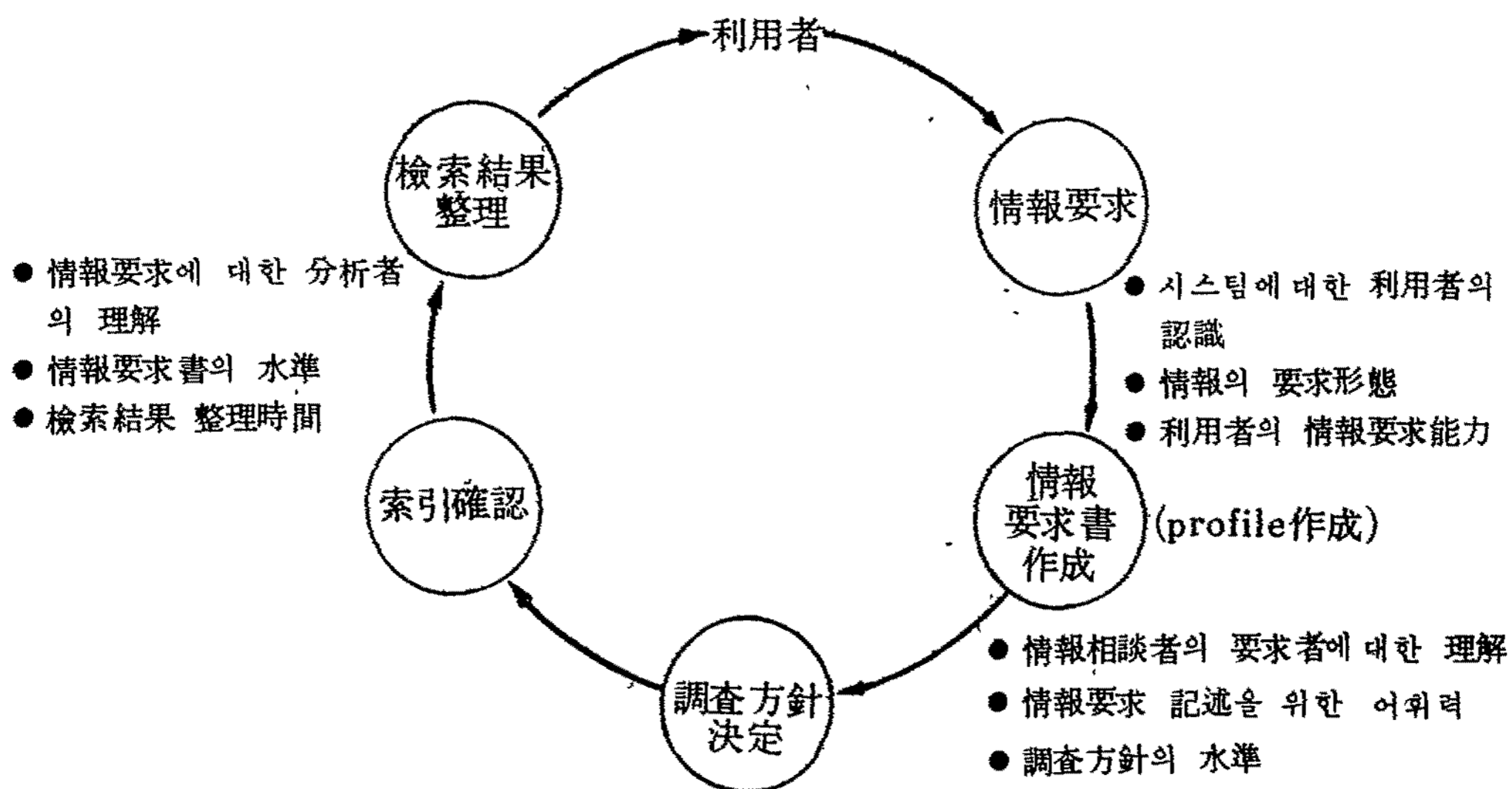


그림 5. 遡及調査시스템의 Work Flow

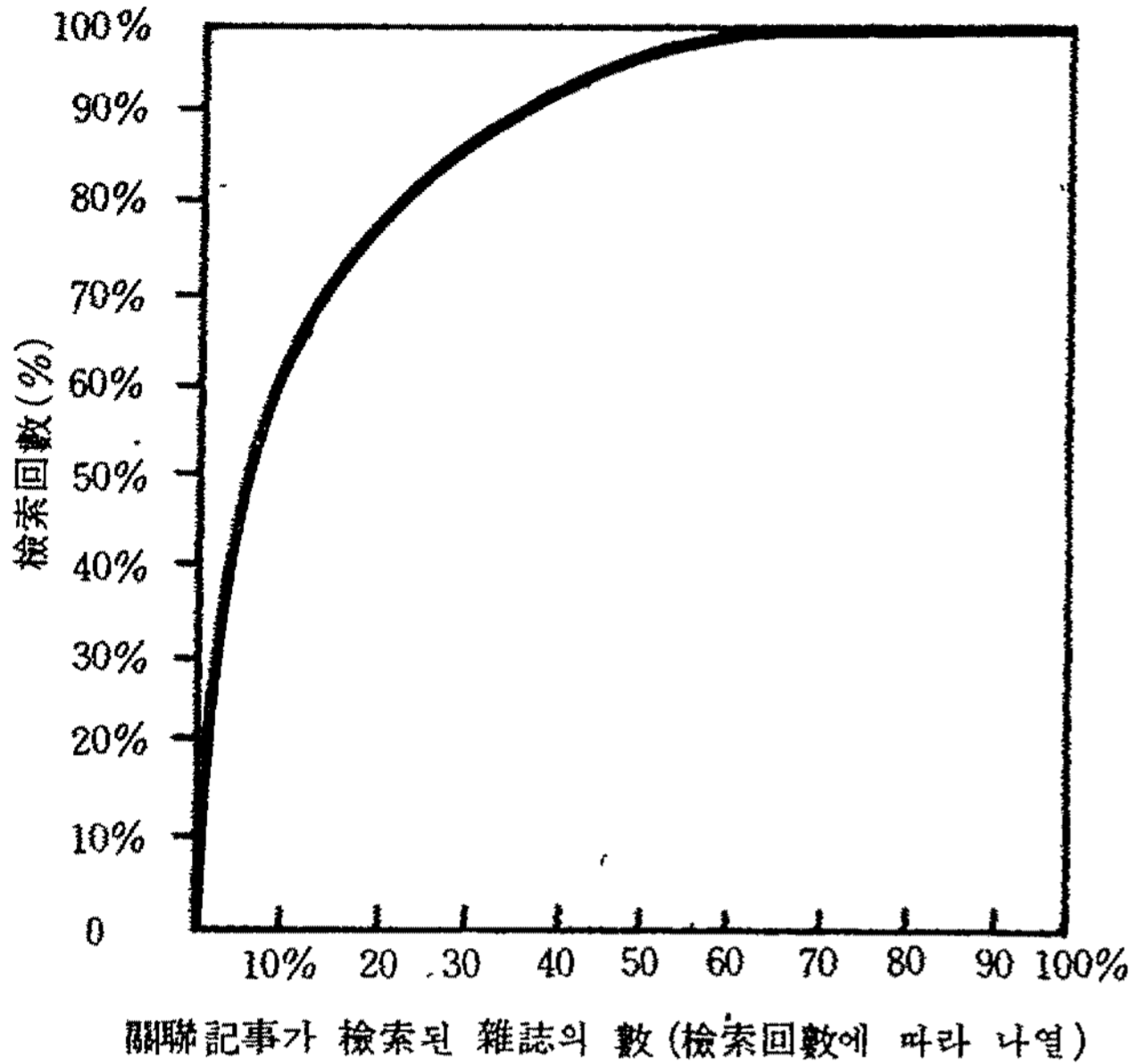


그림 6. 檢索回數對 雜誌의 數

을 통해 각 項目에 대해 投入費用과 利用頻度を 對比시킬 수 있을 것이다. 이와 같은 方法은 시스템의 檢索結果로 나타나는 雜誌에서의 각 項目의 出現回數와 같은 統計資料가 있어야 할 것이다. 이들 統計를 이용하여 檢索에 사용된 雜誌數와 檢索回數를 對比할 수 있다.

그림 6은 對比結果를 圖示하고 있다.

100%檢索을 위해서는 100%의 雜誌利用이 필요하지만 그림 1을 통해 비교적 적은 量의 雜誌가 多數의 檢索에 이용되었음을 알 수 있다. 索引된 雜誌中 30%가 전체 檢索回數의 90%에 이용되었다는 사실은 費用效果面에서 큰 의미를 지니고 있다. 즉 시스템전체의 效率減殺를 極小化시키고 投入費用을 크게 節減시킬 수 있을 것이다. 이러한 研究은 시스템運用 이전에 실시되어야 할 것이다. 그러나 시스템運用 이전에 研究評價한다는 것은 사실 시스템이 운용되고 있을 때 보다 더 어렵다.

現在 이와 같은 豫測評價方法이 몇가지 있다. 즉 Citation Counting을 통한 方法, 圖書館利用分析, 標本利用者(Sample User)를 통해 蒐集한 主題分析方法 등이 이에 속한다.

效率적인 資料選定政策의 例示를 위해 Wiederkehr은 다음과 같은 모델을 設定하였다. 어떤 特定主題에 대해 特定期間동안에 發刊된 N卷의 文獻이 있다. 이 N卷내에, 어떤 特定利用者그룹

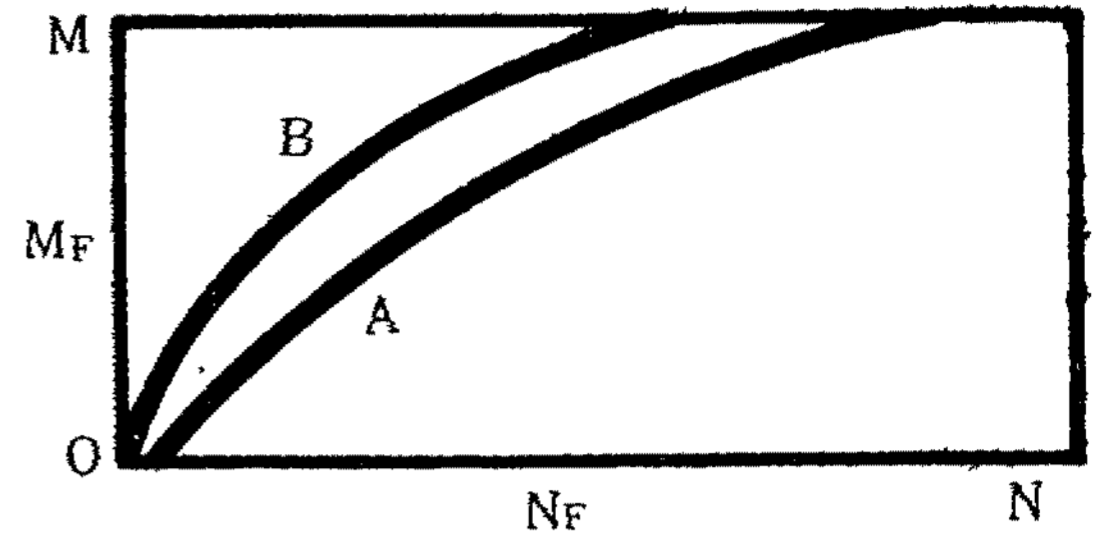


그림 7. Acquisition characteristic curves

에 대해 그들의 情報需要를 만족시켜 주는 文獻이 M卷이 있다. 資料選擇은 전체 N卷 중에서 M卷만을 對象으로 시행되어야 할 것이다. 그러나 실제로는 M卷을 정확히 찾아낼 수 없다. 왜냐하면 시스템에서의 모든 需要를 정확히 豫測할 수 없기 때문이다. 한편 N卷을 모두 구입한다면 有用한 文獻 M卷을 100% 蒐集할 수 있겠지만, (N-M)만큼 불필요한 文獻이 동시에 蒐集되므로 이것은 非經濟的이며 非能率的인 方法이다. 最善의 方法은 有用한 文獻 M을 최대로 구입하고 全体文獻數 N을 最大化하는 것이다. 즉 (N-M)을 極小化시키는 것이다. N_F 를 蒐集處理된 全体文獻의 一部라 하고 M_F 를 蒐集處理된 有用文獻의 一部라 하자. Wiederkehr은 M과 N의 關係를 "Acquisition Characteristic Curve"라고 定義했다. 그림 7은 이 關係의 2가지 형태를 圖示한 것이다.

曲線 B는 曲線 A보다 選擇政策의 效率이 큰 狀態를 表示하고 있다. 曲線上에서 M은 極大化되고 N은 極小化되었기 때문이다. M은 실제 利用記錄을 통해서만 定義할 수 있다. Citation Counts, 圖書貸借記錄 등과 같은 利用形態의 研究은 有用한 資料 M을 包容하는 N의 範疇를 豫測하는데 도움을 줄 것이다.

5. 索引

索引政策과 過程에 있어 經濟的 主眼點은 아래 몇가지 사항에 集中되어야 할 것이다.

- 1) 文獻索引에 所要되는 平均時間
- 2) 索引에 適用시키는 網羅性的의 水準(各 項目에 割當되는 平均索引語數)
- 3) 索引作成者의 專門性

4) 索引改定作業의 必要性

索引政策에 있어 가장 어려운 점은 網羅性的 適正水準을 決定하는 것이다. 즉 索引에 있어 얼마나 많은 索引項目을 使用할 것인가를 決定하는 문제이다. 網羅性이 높아질수록 시스템의 再現率도 높아지지만, 反面에 正確率은 減退된다. 文獻, 索引用語, 情報要求 등의 周邊環境에 따라 그에 부합되는 適正網羅水準이 存在할 것이다.

索引에 있어 費用效果分析은 이와 같은 適正網羅水準을 찾아내는 作業이다. 例를 들어 1主題當 平均 15個의 索引項目으로 어떤 特定文獻을 索引하고 50個의 情報要求를 이용하여 이 檢索시스템을 試驗한다고 하자. 適當한 調查方法을 利用할 때 시스템은 特定情報要求에 대해 73%의 再現率을 나타내고 있다. 情報要求, 調查方法, 索引語 등이 一定한 索引實驗에서 網羅率을 平均索引項目 20個로 適用시킬 때 再現率은 90%로 나타남을 볼 수 있다. 그러나 再現率을 95%로 증가시키기 위해서는 記事當 索引項目을 35個 정도로 網羅水準을 높일 필요가 있다. 이러한 條件下에서는 平均적으로 索引項目 20個를 最適水準으로 看做할 수 있다. 索引項目이 20個가 넘은 후에는 收益遞減現象이 나타나기 때문에 再現率을 全般的으로 5%로 증가시키기 위해서는 網羅率을 索引項目 35個 정도로 증가시켜야 할 것이다. 이와 같은 研究를 통해 再現率과 網羅率을 對比하여 그림 7과 같이 나타낼 수 있다. 이 도표를 통해 索引網羅性的 損益分岐點을 찾아낼 수 있다. 만일 情報利用者の 典型的인 情報要求形態를 갖춘 標本情報要求를 蒐集할 수 있고 이들 情報要求와 관련되는 文獻들을 確認

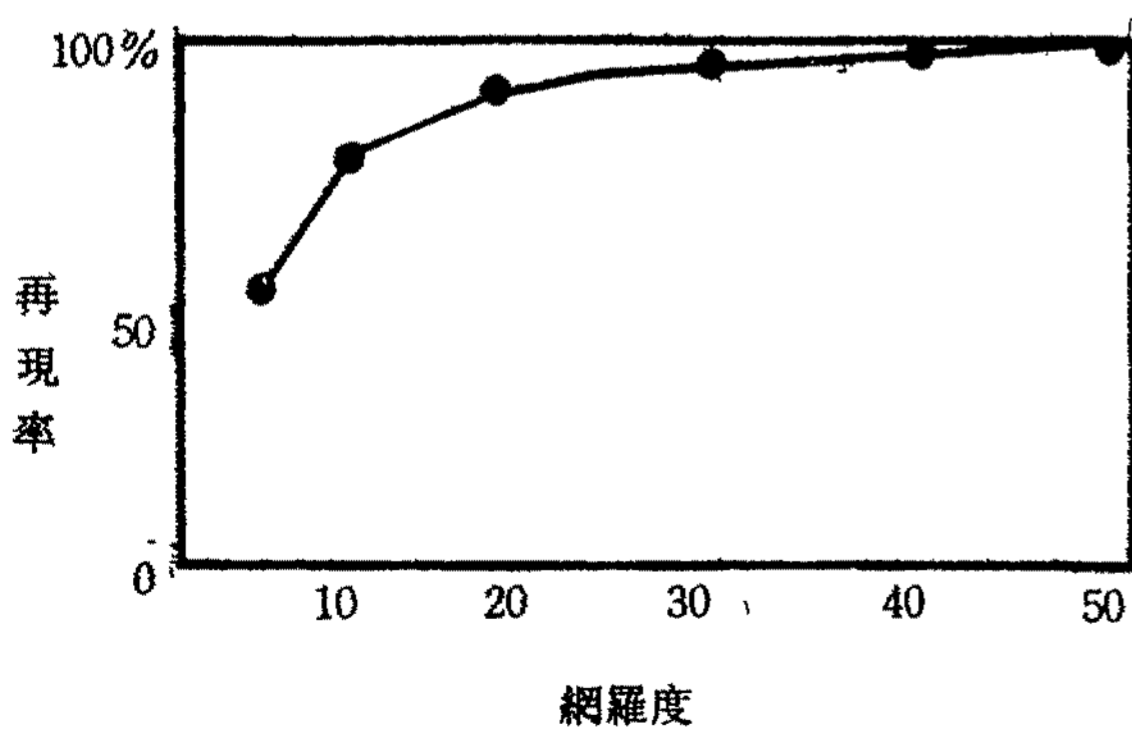


그림 8. 再現率 對 網羅率

해 낼 수만 있다면 이와 같은 費用效果分析은 시스템設計 初期에 쉽사리 適用시킬 수 있을 것이다.

우리는 또한 同一한 문제에 대해 索引作業에 所要되는 時間을 연구점토함으로써 相異한 索引作業들을 分類해 낼 수 있다. 索引作業時間이 증가하면 할수록 索引項目의 數도 증가한다.

추가되는 索引項目數는 비교적 짧은 時間內에 決定되는 것 같다(즉 最初 10分內에 20個項目, 다음 10分 동안에 5個의 索引項目, 최종 10分間에 2개의 索引項目이 추가된다).

有用項目의 90%가 索引所要時間의 最初 10分 안에 處理되고, 나머지 10%의 有用性 여부 판단을 위해 30分이 所要된다면 10分이란 時間은 最適平均索引時間이 될 것이다. 最適網羅水準을 決定하는 방법과 유사한 실험을 통해 索引에 所要되는 時間의 費用效果分析을 시행할 수 있을 것이다. 一般的으로 索引時間이 增加하면 處理되는 項目의 數도 增加되기를 기대해 왔다. 情報檢索시스템의 豫想正確度에 대한 索引時間이나 網羅度의 효과를 測定하기는 매우 어렵다. 그러나 이것은 費用效果分析을 위해서는 절대 필요한 要素이다. 이러한 測定作業은 無作爲 標本推出作業에 의해 修行될 수 있다. 索引에 있어 費用效果分析을 위해서는 文獻의 大量索引과 同一主題에 대해 文獻의 無作爲推出索引이 並行되어야 한다. 索引所要時間을 고려함에 있어 所要時間은 索引의 平均網羅率뿐 아니라 색인의 正確度에도 영향을 미친다는 사실을 기억해야 할 것이다. 索引을 서두르면 서두를수록 索引의 오류는 쉽게 발생한다.

索引에 있어 오류는 2가지 型態로 나타난다.

- 1) 필히 採擇되어야 할 項目의 누락
- 2) 不適合한 項目의 採擇

수시로 發生하는 前者의 경우는 再現率 下落의 原因이 된다. 後者の 경우는 再現率과 正確率 모두의 下落원인이 된다. 索引所要時間에 대한 費用效果分析은 正確率 뿐만 아니라 網羅性에 대한 效果까지도 감안해야 할 것이다. 덧붙여 索引上的 오류는 檢索시스템에서의 再現率에 더 큰 영향을 끼친다는 데 주의하여야 할 것이다.

例를 들어 어떤 特定文獻의 索引에 있어 索引項目 A가 採擇되어야 할 文獻의 90%에 索引項目 B가, B가 採擇되어야 할 文獻의 85%에 索引項目 C가, C가 採擇되어야 할 文獻의 75%에 C가 採擇되었다고 하자. 우리가 項目 A에 대해서만 檢索하였을 때는 再現率이 90%가 된다. 項目 A와 B에 대해 檢索할 때는 再現率은 76.5% ($90\% \times 85\%$)로 떨어지게 된다. 項目 A, B, C에 대해 檢索한다면 再現率은 더욱 떨어져 54.7% ($90\% \times 85\% \times 75\%$)에 지나지 않게 된다. 費用效果의 觀點에서 索引의 正確性에 대한 索引時間의 영향을 고려할 때 檢索性能의 倍增效果를 감안해야 할 것이다. 이것은 우리가 시스템의 어휘를 통해 項目의 平均的인 正確度を 推定하고 이것을 調査의 組合水準에 連結시킬 수 있어야 한다는 의미이다.

索引에 있어 오류의 發生을 全적으로 방지한다는 것은 어려운 일이다. 따라서 모든 索引作業에는 오류를 수정할 수 있는 修正作業이 뒤따라야 할 것이다. 索引修正作業은 費用效果分析面에서 다음과 같은 根據가 있어야 한다.

- 1) 未修正索引에서의 오류의 數
- 2) 修正作業에서 修正된 오류의 數
- 3) 檢索成果에 대한 索引오류의 推定效果
- 4) 修正費用

比較적 간단한 實驗을 통해 推定修正比率과 修正費用을 알아낼 수 있다.

索引作業에 있어 고려해야 할 사항 중 索引을 擔當할 人的 資源은 매우 중요하다. 만일 教育水準이 별로 높지 않은 要員을 索引擔當者로 고용한다면 索引費用은 크게 節減될 것이다. 몇 개의 대규모 회사는 이같은 作業에 크게 성공하였다. 즉 大卒以上の 索引要員을 점차 非大卒者로 교체하였다. 良好한 索引者가 되기 위해서는 正規教育을 받아야 한다는 說이 있으나, 이것은 統計적으로 確定된 것은 아니다. 索引者가 갖춰야 할 能熟度는 다음의 몇가지 사항으로 決定된다.

- 1) 취급하고 있는 主題의 難易度
- 2) 使用하고 있는 索引語
- 3) 索引의 網羅性和 特殊性, 技術的인 面의 索引이 필요할수록 關係分野에 能熟해야 한다.
- 4) 索引作業에 利用되는 각종 도구
- 5) 索引訓練프로그램

索引作業에 대한 費用效果分析은 人的資源에 관한 研究가 先行되지 않고는 完全히 實行될 수 없다. 따라서 우리는 먼저 各계各층의 人的資源을 對象으로 索引作業을 실시케 하여 標準索引과 비교하여 索引의 內容과 分野에 따라 適正 人的資源을 分類해 내야 할 것이다(계속).